



TITLE:

ループリックを活用した学生と教員の評価のズレに関する学生の振り返りの分析 - PBLのパフォーマンス評価における学生の自己評価の変容に焦点を当てて -

AUTHOR(S):

斎藤, 有吾; 小野, 和宏; 松下, 佳代

---

CITATION:

斎藤, 有吾 ...[et al]. ループリックを活用した学生と教員の評価のズレに関する学生の振り返りの分析 - PBLのパフォーマンス評価における学生の自己評価の変容に焦点を当てて -. 大学教育学会誌 2017, 39(2): 48-57

ISSUE DATE:

2017-11

URL:

<http://hdl.handle.net/2433/243350>

RIGHT:

発行元の許可を得て掲載しています。

〈研究論文〉

# ループリックを活用した学生と教員の評価のズレに関する 学生の振り返りの分析

— PBLのパフォーマンス評価における学生の自己評価の変容に焦点を当てて —

斎 藤 有 吾・小 野 和 宏・松 下 佳 代  
(京都大学) (新潟大学) (京都大学)

## Analysis of Students' Reflections on the Differences between Student Self-Assessment and Teacher Assessment through the Use of a Rubric: Focus on Student Self-Assessment Change in PBL Performance Assessment

Yugo Saito, Kazuhiro Ono, Kayo Matsushita  
(Kyoto University) (Niigata University) (Kyoto University)

A current debate surrounding educational reform focuses on self-assessment abilities. In this research, during the performance assessment of problem-solving abilities, students were asked to assess themselves. We then attempted to promote students' reflection on the reasons behind the differences between self- and teacher assessment. Using both quantitative and qualitative data, we examined how reflection affected students' self-assessment. The results indicated that students increased their awareness of the meaning of the rubric descriptors and what was required by the tasks. Finally, we discovered that students were likely to determine their own weaknesses based on the differences from the teacher assessment and create an in-depth reflection that leads to further learning. This type of "assessment as learning" could also indicate a promising direction for utilizing rubrics in students' learning.

〔キーワード：PBL，改良版トリプルジャンプ，ループリック，自己評価，直接評価，振り返り〕

### 1. 問題と目的

#### (1) 問題意識

##### 1) ループリックの学びへの活用

昨今，アクティブラーニングの急速な普及とともに，パフォーマンス評価と評価基準表であるループリックに注目が集まっている。ループリックは複数の評価者間で評価の一貫性を確保するために行う調整作業（モデレーション）のツールとしての機能や，学習の質を量に変換するという機能を持つ（松下，2012）。そのためループリックはパフォーマンス評価における複数評価者間の評

価の信頼性を検討したり（例えば，松下他，2013），「公平で客観的かつ厳格な成績評価」を担保しようとしていたりする試みなど（例えば，沖，2014）で利用されている。

しかし松下（2012）が指摘するように，本来パフォーマンス評価やループリックは心理測定学的パラダイムではなく，構成主義や状況論などを基盤としており，形成的評価によって教育改善・指導，学生の成長を促すことが主目的となる。山田・森・毛利・岩崎・田中（2015）ではそのような観点から，ループリックを学びに活用するための知見を，多くの研究事例のレビューを通して提示している。そのような研究のなかでも寺嶋・林（2006）は，学生が自ら学習やその評価を一体化さ

せつつ、自己学習力を向上させる必要があるという問題意識から、ループリックを事前に学生に提示し、それをもとに学生自身に振り返らせるという実践を行っている。そしてその効果をアンケート調査から検討し、自分自身の課題や学習方法の重要性を認識させ、目標を意識化させる効果があることを明らかにしている。この研究はループリックを学びに活用し、特に学生の自己評価能力の向上を目指した研究として位置づけることができる。

## 2) 自己評価能力への着目

田中(2008)によると、自己評価能力とは、学習者が自分で自分の人となりや学習の状態を評価し、それによって得た情報によって自分を確認し今後の学習や行動を調整する能力のことである。また、情報社会の進展と生涯学習社会の到来を背景として、生涯にわたり学習するための自己学習能力の形成が求められるようになってきており、自己学習能力は自己評価能力に裏打ちされない限り十全に機能しないため、その形成の重要性が問われていると述べている。

事実、昨今の教育改革のキー概念である「資質・能力」では複数のモデルが提案されているが、自己評価あるいはメタ認知はそれらのモデルに共通して見出すことができる。松下(2016)は国内外で議論されており影響力をもっている複数の「資質・能力」概念を①<育成すべき資質・能力に包摂される個人の属性>に着目したもの(3次元構造)と②<資質・能力を育てる関係性>に着目したもの(3軸構造)に整理し、どちらにおいても「省察性(reflectiveness)」(自らの学びについて省察的・メタ的であること)を求めていることを明らかにしている。なお、典型例として、①ではカリキュラム・リデザイン・センター(CCR)の4次元の教育(Fadel et al., 2015), ②ではOECD-DeSeCoのキー・コンピテンシー(Rychen & Salganik, 2003)が挙げられている。

そして、松下(2016)はこれらを統合的に捉える枠組みとして「3・3・1」モデルを提唱している。「1」に該当する省察性には、メタ学習(メタ認知と成長的マインドセットが含まれる)あるいは学ぶことの学習に加え、批判的なスタンス、差異・矛盾への対応などの意味合いも含まれている。これらを機能させるためには、先述の田中(2008)が指摘するように、自己評価能力の裏打ちが必要になるだろう。このように、現在の教育改革において自己評価能力の形成が求められていることがわかる。

また、そのような教育改革の下で普及が進むアクティ

ブラーニングにおいても、学習に対する自己評価とその結果にもとづいた学習の自己調整が求められている(松下・石井, 2016)。例えば、須長(2010)はアクティブラーニングにおいて学習者は単に個別にそれぞれのセッションを実施するだけに留まらず、全体の関係性の中で個々のセッションの意義付けを行うことが重要であり、そしてそのような全体調整には、一連のプロセスを完遂するための観察・遂行制御・反省等を行う自己調整が必要であることを指摘している。また、自己調整学習方略のリストには自己評価が含まれており(Zimmerman, 2001; 伊藤, 2012), 主体的に学習を進める際に、適宜ふり返って計画の遂行状況や理解度を自己評価できなければ、自分の力で学習を進めたり次の学習行動につなげたりすることはできないため、やはりアクティブラーニングにおいてその重要性を指摘できる(中山, 2013)。

このように、昨今の社会的要請や現在の教育改革を背景として、またそれの下で普及が進むアクティブラーニングにおいても、学習の自己調整に関わる重要な概念として自己評価能力は注目されていることがわかる。

## 3) 学生を評価主体として育成するためのループリックの活用

特に高等教育分野において、学生が自分の状態をもっともらしく把握できるのかどうかを実証的に検討したもののとして、斎藤・小野・松下(2017)がある。そこでは、アカデミック・ライティング能力を共通のループリックを用いて教員による評価と学生の自己評価から捉えているが、教員による評価と学生の自己評価では、学生の自己評価は教員による評価と比較してかなり甘めであり、さらにそれがほぼ無相関であったことを報告している。すなわち、共通の評価基準を用いたとしても、エキスパートの教員による評価とノービスの学生の自己評価は絶対的にも相対的にも大きくズレることを明らかにしている。

なお、松下・小野・高橋(2013)では同様のパフォーマンス評価において、複数名の教員同士の評価による信頼性を一般化可能性理論によって検討し、一定の信頼性が得られていることを報告しているが、これをズレという観点から見ると、エキスパート同士の評価であっても、やはりある程度のズレは生じることがわかる。しかし、斎藤他(2017)で見られたような顕著なズレではない。

つまり、ある教員による評価を、絶対的に正確な評価とみなす(測定論でいえば、誤差が全く含まれておらず、真値と同じものとみなす)ことは危険ではあるが、それでもやはり、教員同士ではズレが小さく、教員と学

生とではズレが顕著に大きいということから、教員による評価を、エキスパートの鑑識眼を通したより妥当な評価として位置づけることができるだろう。そのように位置付けた場合、教員による評価とのズレが大きい学生は、自分の学習の状態をもっともらしく把握することができていない、ということになる。

特にこれは、高等教育において涵養が目指されている高次の統合的な能力を自己評価する際に重要になる。Sadler (2010) が指摘するように、複雑な能力の表れであり、標準化できないような成果物に対しては、様々な観点からその質を総体的に判断する必要がある。その際、「批評せよ」と「説明せよ」など、課題で指定された応答ジャンル (response genre) によって求められている回答が異なるということや、「一貫性」や「エレガントさ」といったルーブリックなどの基準で用いられている言葉の意味を理解して (あるいは評価者同士で共有して) いなければならない。つまり、高次の統合的な能力を評価するという行為には、応答ジャンル、基準、質といった概念とその相互関係の理解を前提とした専門的技術が必要である。山田・斎藤 (2017) はそのような背景から、成績評定に関しては学生の評価を教員の解釈を通して限定的に使用する必要性を示しつつも、そのような評価に学生が参加することで、学生を評価主体として育てていく可能性を論じている。Sadler (2010) や山田・斎藤 (2017) の指摘は学生同士のピア評価という文脈でなされたものであるが、自己評価においても同様の指摘ができるだろう。

さて、学生を評価主体として育成する際に、ルーブリックは強力なツールとして位置づけることができる。ある課題に対する成果物から、何かしらの能力を評価するために教員が作成したルーブリックは、その課題の応答ジャンルや質を判断する基準を明文化したものとなる。学生はそれをもとに自身の成果物から自身の能力を評価するという経験をする中で、どのような視点で質を判断すべきなのかということを学ぶことができる。ルーブリックを事前に提示し、学生に自己評価させる実践を行った、先述の寺嶋・林 (2006) の研究において、評価の視点に気を配ることや、自分の能力を客観的にみることにできていたとアンケートに回答する学生が多かったこともそれを支持している。ただし、教員による評価と学生の自己評価とのズレは検討していないため、本当に学生が適切な自己評価を行えていたかどうかは不明である。

また、斎藤他 (2017) が示したように、ルーブリックを事前に提示したとしてもなお、教員と学生の評価に

は大きなズレが生じる。ここには、Sadler (2010) がいう応答ジャンル、基準、質といった概念とその相互関係の理解の程度が影響していると考えられる。言い換えれば、エキスパートとノービスの鑑識眼の違いがこのようなズレを引き起こしていると考えられる。また、能力の低い人が自身の能力を過剰に評価する傾向にあり、能力の高い人は逆の傾向にあるという認知バイアス (ダニング=クルーガー効果) の影響も考えられよう (Kruger & Dunning, 1999)。そこで、ただ自己評価をさせて「教員による評価とズレている」と認識させるだけではなく、「なぜ教員による評価とズレたのか」に関して意識的に振り返りをさせることで、学生は教員の鑑識眼との隔たりを認識するだけでなく、評価に関わる概念の理解を深め、それに伴って自己評価能力を高めることができるのではないだろうか。このようにルーブリックを、単に学習の質を量に変換するためのツールとして用いるのではなく、学生を評価主体として育成するために活用することは、形成的評価によって教育改善・指導、学生の成長を促すという本来的な目的に合致するものといえる。

## (2) 研究目的

以上のような問題意識から本研究では、あるPBL科目における問題解決能力を評価するためのパフォーマンス評価を事例として、ルーブリックを用いて学生に自己評価させるだけではなく、学生の自己評価と教員による評価とのズレを学生に提示し、「なぜズレたのか」の振り返りを促す実践が、学生の自己評価にどのような影響を与えるのかを、ズレの度合いと振り返りの記述内容という量的・質的データの双方から検討することを目的とする。

## 2. 方法

### (1) 新潟大学歯学部におけるPBL授業におけるパフォーマンス評価

本研究において事例として取り上げるのは、新潟大学歯学部口腔生命福祉学科において2年次の前期・後期にわたり開講されているPBL科目群におけるパフォーマンス評価である。この科目は、学生自身が口腔に関わるテーマから問題を見出し、仮説を立て、学習課題を設定し、授業時間外で学習課題を調査・自習したり、学習課題に関連したセミナーで新たに得たりした知識をもとに問題解決に取り組むというものである。

ここでは、PBLにおける問題分析・問題解決能力、自己学習能力を評価するために考案された、「改良版トリプルジャンプ」というパフォーマンス評価を実施して



前期	後期
「わたしって、ダメな歯学部生?！」	「わたし、困っています」
<p>あなたは新潟大学歯学部2年生です。4月から専門科目の授業が始まり、解剖学や生理学を学んでいます。『PBLチュートリアル』という新しい学習方法にまだ慣れず、また学習内容も急に難しくなったように感じられ、不安を抱えながら悪戦苦闘の毎日です。</p> <p>そんなある日、全学のサークルで知り合った友人の佐藤 彰くん(工学部3年生)が、左の頬を腫らして近づいてきました。彼は3日前に、近くの歯科医院で、下顎の左側智歯を抜去したとのことですが、いまだに左側の下唇に麻酔がかかっているような感じが残り、ご飯粒がついていてもわからないといいます。また、抜歯してから口を大きく開けられなくなり、飲み込むときにノドの左側に痛みもあるとのこと、食事がとりにくいと訴えています。</p> <p>あなたが歯学部の学生であることから、なんでこんなことが起きているのか説明してほしいと頼まれましたが、あなたはどうか答えてよいかかわからず黙っていました。彼は不安そうな顔をして、「次のサークルのときでいいから」と言って別の友人のところに去っていき、彼の後ろ姿を見ながら、将来、歯科医療従事者になるものとして、せめて「大変だね」のひと言でもかけてあげればよかったと、あなたはちょっと後悔しました。</p>	<p>あなたは新潟大学医学歯学総合病院の歯科衛生士です。</p> <p>今日は担当患者：高橋勇蔵(67歳・男性)の2回目の診療日です。高橋勇蔵は中等度の歯周病があり、初回は歯周検査と病状説明を行いました。</p> <p>あなた：高橋さん、お口の具合はいかがですか。前回、タバコをやめるようお話ししましたが、禁煙されましたか。</p> <p>高 橋：してないよ。私はね、タバコをやめるくらいなら死んだ方がましだと思っているよ。この前、国から送られてきたアンケートにも「生きがいはタバコを吸うこと」と書いてくれた。相変わらず1日40本は吸っている。歯科に来て、なんでタバコをやめるよう言われなきゃならんのだね。</p> <p>あなた：でも、高橋さんは糖尿病もあるし、やめた方がいいと思いますが・・・</p> <p>高 橋：糖尿病は関係なかろう。ここは歯科だろう。おや、内科と間違えたかな。</p> <p>あなた：歯科ですけど・・・とにかく、前回言ったことと同じことを言いますが、まずはタバコをやめてください、いいですね。</p> <p>高 橋：ああ、わかった。あんたはタバコが嫌いだな。</p>

図1 改良版トリプルジャンプ用シナリオ（前期・後期）

いる。問題解決能力を評価するために用いられた課題とルーブリックは歯学と教育評価の専門家によって内容的妥当性、表面的妥当性が確認され、複数名の評価者間の評価の整合性の観点から信頼性が確認されている。詳細は小野・松下・斎藤(2014)を参照されたい。

改良版トリプルジャンプは、PBLにおける学生の学習成果を評価するために開発された方法であり、以下の3つのステップからなる。ステップ1では、提示されたシナリオ(図1)から問題を見つけだし、いったん既有知識で解決策を立案した上で、調べるべき学習課題を設定し、その過程をワークシートに記述する(60分間)。ステップ2では、学習課題を調査・学習し、その結果をもとに当初の解決策を検討して、最終的な解決策を提案し、その過程をワークシートに記述する(1週間)。最後のステップ3では、シナリオにもとづく場面設定で、教員を相手にしたロールプレイによって、解決策の実行を実演する(15分間。ただし教員からのフィードバックを含む)。ワークシートによる筆記課題とロールプレイという実演課題を組み合わせ、2つの異なるタイプのルーブリック(ステップ1・2は様々な評価課題に対応できる一般的ルーブリック、ステップ3は課題特殊のルーブリック)を用いたパフォーマンス評価である。

なお、ステップ1・2においてのみ、教員が使用するルーブリックを学生に事前に示し、自己評価をさせている。本研究では教員による評価と学生の自己評価とのズレやその振り返りの内容を検討することを目的としているため、ステップ1・2の評価を検討対象とする。

## (2) 自己評価とその振り返り

2014～2016年度において、改良版トリプルジャンプではすべて同じ課題・ルーブリックによって同じ手続きがとられている。ただし2015年度から、本研究のような問題意識によって、学生に自己評価をさせるだけでな

く、教員による評価とのズレに関して振り返りをさせる活動を新たに加えた。具体的には、ステップ1・2のルーブリック(6観点)と、各観点における教員による評価と学生の自己評価を同時に返却し、ルーブリックを参照しながら2つの評価のズレをすぐに確認できるようなプリントを作成した(表1)。そのプリントでは、教員による評価と自己評価をもとに、「現在の到達状況」「教員による評価と自己評価がズレた理由(ズレがない場合は書く必要はない)」「今後観点ごとの力やそれらを統合した問題解決能力をどうしたら伸ばすことができるかの考え」を自由記述によって回答を求めるようになっている。なお、プリントに記載されている教員による評価は、歯学を専門とする著者のうちの一人によるものである。

まとめると、2014年度と2015・2016年度との間にはそのような振り返りの活動を行ったかどうかの違いがある。PBL科目は前期と後期に開講されており、改良版トリプルジャンプもそれぞれの期末に行われているため、①2014年度と2015・2016年度の前期と後期において教員による評価と学生の自己評価のズレがどのように変化したのかの量的な比較と、②2015・2016年度において前期と後期で振り返りの記述内容がどのように変化したのかの質的な比較が可能である。本研究では特に、「教員による評価と自己評価がズレた理由」への記述内容に注目する。これらの比較を通して、このような活動が自己評価にどのような影響を与えるのかに関する示唆が得られると期待できる。

## (3) 変数化の手続き

「問題発見」「解決策の着想」「学習課題の設定」「学習結果とリソース」「解決策の検討」「最終解決策の提案」の6つの観点からなるステップ1・2のルーブリック(表1)を用いて、著者のうちの一人である教員1名に

表1 自己評価のリフレクションを促すためのプリント（抜粋）

ワークシートの評価基準						
観点	1. 問題発見	2. 解決策の着想	3. 学習課題の設定	4. 学習結果とリソース	5. 解決策の検討	6. 最終解決策の提案
観点的説明	シナリオの事実から問題を見いだす。	解決の目標を定め複数の解決策を立案する。	問題の解決に必要な学習課題を設定する。	信頼できるリソースから学習課題を調査する。	解決策の有効性や実行可能性を検討する。	問題に対して最終的な解決策を提案する。
レベル3	問題を見いだし、シナリオの事実から、推察する原因も含め、問題とした理由を述べている。	いくつかの解決策を立て、これまでの学習や経験とも結びつけて、解決策の立案過程を述べている。	学習課題を的確に設定し、解決策と学習課題の関連から必要性を述べている。	利用可能なさまざまなリソースを駆使し、信頼性に注意して、正しい内容を学習している。	いくつかの解決策を比較検討し、それぞれの有効性や実行可能性を考察している。同時に、解決策の限界にも思いをめぐらしている。	シナリオの状況に適した、妥当な最終解決策を提案している。解決策をより効果的に実行するために追加情報の必要性に気づいている。
レベル2	問題を見いだし、シナリオの事実から、問題とした理由を述べている。	いくつかの解決策を立て、解決策の立案過程を述べている。	学習課題を設定し、解決策と学習課題の関連から必要性を述べているが、重要な学習課題が一部欠如している。	リソースの信頼性に注意して、おおむね正しい内容を学習している。	いくつかの解決策を比較検討し、それぞれの有効性や実行可能性を考察している。	シナリオの状況に適した、妥当な最終解決策を提案している。
レベル1	問題を見いだしているが、問題とした理由の説明は不十分である。	解決策を立てているが、立案過程の説明は不十分である。あるいは解決策が1つのみである。	学習課題が漠然としており、何を学ぶべきか焦点が絞られていない。あるいは必要性の説明が不十分である。	リソースの信頼性についての注意が不十分で、学習内容にいくつかの誤りが含まれている。	解決策の検討は不十分である。あるいは複数の解決策について比較検討していない。	最終解決策の提案にいたっていない。あるいは解決策、学習結果、結論の間に矛盾や飛躍がある。
レベル0	レベル1を満たさない場合はゼロを割り当てること。					

学生氏名：〇〇

教員評価結果	2	2	3	2	1	2
自己評価結果	1	1	3	1	3	1

上記の評価結果で、あなたの自己評価が教員評価と異なった理由を観点ごとに考えてください（評価結果にズレがない観点については書く必要はありません）。なお、教員評価の結果に納得がいけない場合は、自己評価結果の方が妥当とする理由を説明してください。

5: 解決策は複数と比較などして十分に検討できていると思っていたが、今考えてみると、有効性や実行可能性をどのように言及したら「十分な検討」たりえるのかをよくわかっていなかった。特に「有効性」はどのような視点で検討したらよいのか、再考してみたい。

※ワークシートの評価基準は、改良版トリプルジャンプ1・2の評価基準に該当する。また、評価結果や記述内容は架空のものである。

よる評価と、学生の自己評価を、レベルに応じて得点に換算した（0～3点）。また、このループリックはPBLにおける統合的な問題解決能力を評価するためのものである。そのため、6観点の平均点は「PBLにおける問題解決能力」と操作的に定義する。

そして、本研究では教員による評価をエキスパートの鑑識眼を通したものとして、学生の能力を捉える際のより妥当なものとして位置づけ、そこから学生の自己評価を減じた差分をズレとする。そのように変数化したズレが前期と後期でどのように変容するのかを、振り返りがあった場合となかった場合とで比較することで、振り返りの影響を検討する。

なお、評価のズレの度合いを検討するために、ズレの絶対値やズレの分散をとってその変容を比較する手法も考えられる。しかしそのような手法では「評価の甘さ／辛さ」などの「どのようにズレるのか」の情報が損なわれてしまうため、本研究では単純な差分を使用する。ただし差分のみの情報では、例えば全体のズレの平均値が0であったとしても、教員と学生の評価が完全に一致している場合と、甘めの評価や辛めの評価をする学生が混在しているためにそれらが相殺されている場合などがあり得るため、平均値のみで解釈することは危険である。そこで、全体としてズレがどのようになったのかを検討

する場合には、全体のズレの平均に加えて、そのバラつきも併記し、考察の対象とする。

#### (4) 調査対象者及び手続き

調査対象者は、2014～2016年度においてPBL科目群を前期・後期に履修した2年次生60名（年度順に、18名、22名、20名）である。データの収集・使用にあたり新潟大学歯学部倫理委員会の承認（承認番号26-R54-03-11）を得たうえで、調査対象となる全ての学生から同意を得た。

### 3. 結果と考察

#### (1) 前期から後期にかけての評価のズレの度合いの変容の比較

前期と後期の教員による評価と学生の自己評価の要約統計量を2014年度（振り返りなし）と2015・2016年度（振り返りあり）別に表2に示す。ズレの度合いの平均値が正の値であれば、教員による評価と比べ相対的に学生の自己評価が辛く、負の値であれば自己評価が甘いことになる。観点ごとに若干傾向は異なるが、概ねどちらも前期から後期にかけて、ズレの度合いの平均値の絶対値は0に近づいている。ただし、標準偏差から見るバラつき具合は観点によって一貫した傾向は見られず、後期のほうがバラつきが大きいものもある。「全観点平均」

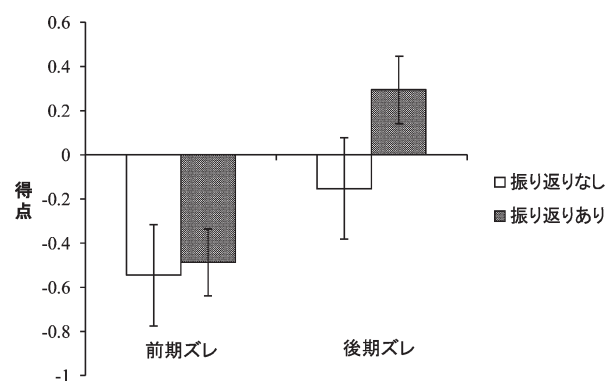
表2 前期と後期の教員による評価と学生の自己評価の要約統計量

観点	時期	2014年度 (n=18) 振り返りなし						2015・2016年度 (n=42) 振り返りあり					
		教員による評価		学生の自己評価		ズレの度合い		教員による評価		学生の自己評価		ズレの度合い	
		平均値	SD	平均値	SD	平均値	SD	平均値	SD	平均値	SD	平均値	SD
問題発見	前期	1.83	0.62	2.06	0.42	-0.22	0.81	1.79	0.72	2.10	0.58	-0.32	0.93
	後期	1.50	0.51	2.00	0.34	-0.50	0.71	2.24	0.86	1.93	0.41	0.32	0.91
解決策の着想	前期	1.17	0.71	2.00	0.49	-0.83	0.79	1.52	0.59	1.98	0.57	-0.46	0.74
	後期	1.78	0.55	1.67	0.69	0.11	0.90	1.88	0.64	1.71	0.56	0.17	0.80
学習課題の設定	前期	1.33	0.49	1.94	0.64	-0.61	0.98	1.40	0.50	2.10	0.66	-0.71	0.84
	後期	2.00	0.84	1.78	0.43	0.22	0.81	2.20	0.78	1.51	0.60	0.68	0.88
学習結果とリソース	前期	1.56	0.62	2.11	0.32	-0.56	0.70	1.86	0.57	2.02	0.52	-0.15	0.73
	後期	2.22	0.65	1.88	0.33	0.41	0.62	2.22	0.61	1.85	0.61	0.37	0.54
解決策の検討	前期	1.06	0.64	1.78	0.43	-0.72	0.83	1.00	0.66	1.73	0.50	-0.71	0.75
	後期	0.67	0.77	1.50	0.51	-0.83	0.79	1.24	0.73	1.34	0.57	-0.10	0.89
最終解決策の提案	前期	1.72	0.67	2.06	0.42	-0.33	0.69	1.43	0.59	2.07	0.52	-0.63	0.62
	後期	1.61	0.61	1.82	0.53	-0.18	0.81	1.90	0.49	1.63	0.54	0.27	0.59
全観点平均	前期	1.44	0.45	1.99	0.23	-0.55	0.56	1.50	0.39	2.00	0.35	-0.50	0.43
	後期	1.63	0.41	1.78	0.31	-0.15	0.47	1.95	0.44	1.66	0.38	0.28	0.50

に関しては、2014年度（振り返りなし）と2015・2016年度（振り返りあり）のどちらもズレの度合いの平均値が前期から後期にかけて絶対値が0に近づいている。ただし2014年度（振り返りなし）後期は負の値を示しており学生の自己評価のほうが相対的に甘く、2015・2016年度（振り返りあり）後期は正の値を示しており学生の自己評価のほうが相対的に辛いことがうかがえる。また、標準偏差は0.5付近であり、どちらも前期から後期にかけて顕著な変容はない。

これらを総合的に解釈すれば、全体としては、前期は年度に関わらず、学生は教員による評価と比較して甘めに評価する学生が多かったことがうかがえる。そして後期になるとズレの度合いの平均値が絶対値としてはどちらも0に近づいたが、ズレの度合いのバラつきは前期から後期にかけてほぼ変わっていないため、学生の自己評価が教員による評価と一致するようになったというわけではなく、教員よりも甘めに評価する学生と辛めに評価する学生が混在するようになったことがうかがえる。

次に、ズレの度合いが前期から後期にかけてどのように変容したのかに関して全体的な傾向を把握するため、「全観点平均（PBLにおける問題解決能力）のズレ」を従属変数、「振り返りなし／振り返りあり」の群（参加者間）、「前期／後期」の時期（参加者内）を独立変数として、2要因の分散分析（混合計画）を行った。また、そのグラフを図2に示す。その結果、群と時期の有意な交互作用効果が見られた（ $F(1,56)=5.45$ ,  $p<.05$ ,  $\eta_p^2=.09$ ）。これは、前期から後期にかけてのズレの変容の様相が、群間で異なることを示している。それがどのように異なるのかを検討するために、単純主効果の検定（Holm法による調整）を行ったところ、後期における群間の単純主効果が有意であった（ $t(112)=3.26$ ,



※エラーバーは95%信頼区間

図2 PBLにおける問題解決能力のズレの比較

$p<.05$ ,  $g=1.3$ ）。これらの結果から、図2からもわかるように、振り返りあり群のほうが振り返りなし群と比較して、ズレの度合いが前期から後期にかけて顕著な変容を示しており、より学生の自己評価が相対的に辛くなるというものであった。ただし先述のように、それによって教員による評価と一致するようになったわけではない。つまり、振り返りによって後期の自己評価の際に慎重になり、教員よりも相対的に辛めに評価する学生が増え、前期とは逆向きのズレが生じていることが示唆される。

これらのことから、教員による評価をより妥当なものとした場合に、ルーブリックを活用した評価のズレに関する振り返りによって、学生の自己評価が相対的に辛めになる可能性が示唆された。上述のように、逆向きのズレを引き起こすことも考えられるため、自己評価が辛めになるということがすなわち自己評価能力が高まったということを主張することはできないだろう。

このような変容が自己評価能力の形成にどのような意味を持つのかは定かではないため、今回の結果から、振



り返りが自己評価能力に対してポジティブな影響を与えたのか、ネガティブな影響を与えたのかを論じることは現時点では困難である。このように自己評価が辛めになるということが、長期的に見れば自己評価能力の形成に必要なかもしれない。逆に、学生の自己効力感を過度に損なってしまったのかもしれない。では、このような自己評価の変容がどのような意味を持つのだろうか。次に学生の振り返りの記述内容を分析することによって、以上のことに迫りたい。

## (2) 前期から後期にかけての振り返りの記述内容の比較

2015・2016年度は、前期と後期の期末に、教員による評価と学生の自己評価のズレの振り返りの活動を行っている。この活動に参加した42名の学生の、ズレが生じた理由に関する振り返りの自由記述に関してテキストマイニングを行い、前期と後期でどのような違いが見られるのかを検討した。自由記述の分析にはKH Coder (樋口, 2014) を使用した。

表3 前期と後期の振り返りの記述内容の特徴語

前期		後期	
不十分	.346	内容	.273
説明	.275	理由	.255
比較	.245	原因	.250
調べる	.205	必要	.250
書く	.196	過程	.192

「教員による評価と自己評価がズレた理由」に対する自由記述の前期と後期の特徴語を表3に示す(表の数値はJaccardの類似性測定)。これら特徴語は単なる頻出語ではなく、データ全体に比してそれぞれの部において特に高い確率で出現している語であるため、前期と後期の記述を特徴づける語であるといえる(樋口, 2014)。これらの特徴語をキーワードとして、前期と後期それぞれでこのような言葉を用いている学生の代表的な記述例を示す。なお、その記述をした学生の特徴として、自己評価が教員による評価に比べ甘め/辛めだったかどうか、PBLにおける問題解決能力(教員)の得点が全体の平均に比べどうだったのかを付記する。

まず前期は、「不十分」「説明」「比較」など、課題で求められているものや、ループリックの記述語で使用されている言葉の意味に関する言及が目立った。例えば、「解決策の意味がよく分かっていなく、検討[見当]違いだったため」(学生A:自己評価甘め, 平均以下)という、課題で求められているものの意味がよくわかっていなかったという記述が見られた。また、「問題発見については、理由の説明の量が不十分だったことが原因

だった。解決策の着想については、立案過程を述べたつもりでいたが不十分だった」(学生B:自己評価甘め, 平均程度)のように、どのくらい書いていけば「十分」なのかがわからず、ズレてしまったという記述があった。さらに、「解決策の検討については、「比較」をすることなく、情報をさらに集めたただけだったため、「検討や比較」を行えていなかった」(学生C:自己評価甘め, 平均以上)のように、「検討や比較」には何が求められているのかの認識が異なっていたことに気づく記述なども見受けられた。

他にも「あまり比較していないと感じたが、考察する過程で思っていたより検討できていたためだと考えた」(学生D:自己評価辛め, 平均以上)のように、「比較する」ということに関して教員よりも高めに水準を設定し、辛めに自己評価したという学生もいた。このような記述は特に、PBLにおける問題解決能力(教員)が平均よりも高い学生に散見された。このように前期におけるズレの振り返りの活動は、課題で求められているものや、ループリックの記述語の意味などに目を配るきっかけになっているようである。

それに対し、後期は「理由」や「原因」など、自分自身の成果物の内容面において何が足りなかったのか、足りていたのかを具体的に考察する記述が目立った。例えば、自己評価を甘めにした学生の記述として、「自分なりには説明できたつもりであった。しかし、見直してみると、内容はともかく、『～だから』というような、理由を表すような表現でかけていなかったことに気付いた。『～なはずがない』という言い方を並べても理由の説明にはならないので、書き方にも注意したい」(学生E:自己評価やや甘め, 平均以上)のように、自分の成果物に立ち戻って、教員が求める水準を満たすためには具体的にどのように書くべきだったのかを考察している。

また前期において、課題で求められていることの認識のズレに言及していた学生Cは、「自分の中ではシナリオに出てきているキーワードをチェックし、それが関連するものをただ漠然と課題に設定し、断片的に調べてしまっていたのが原因だと思う…(中略)…検討の仕方を勘違いしていたことが一番の原因。自分が設定した学習課題と解決策の着想を見比べ、これで本当に問題は解決するのか、実行できる範囲で考え、限界にも思いめぐらす必要があった」(学生C:自己評価甘め, 平均的)というように、なぜズレてしまったのか、またそれを解消するためには、あるいは次の学習ではどうしたらよいのかということにつながる言及をしている。



他方、自己評価を辛めにした学生の記述として多かったのは「問題発見については、自分が思っていたよりも詳しく問題とした理由が述べられていたからであると考える」(学生F：自己評価辛め、平均以上)というように、前期の学生Dのように、高めに水準を設定し、辛めに自己評価したというものである。前期と異なることは、PBLにおける問題解決能力(教員)が平均的～平均より低い学生にも散見されるようになったということである。

そのような学生の記述の中には、「解決策の着想については、その解決策をたてた理由をうまく言葉にできなかった。そのため『1』という評価にしたが、解決しようと注目した観点は悪くなかったようだ。最終解決策の提案では、調査の量が膨大で、まとめるのが難しかった。そのため、詳しく説明すべきところで内容が薄くなってしまった。しかし、患者との信頼関係を築くポイントや治療にあたって説明する意義、歯科衛生士の役割など理解した上で患者とコミュニケーションがとれたという点では良かったと思う」(学生G：自己評価辛め、平均以下)のように、単に「思ったよりよかった」というのではなく、不安になっていたものの、教員による評価から自己評価の修正を行い、また良かった点を把握しようとしている。

同様に、「自分がインターネットで得た情報が本当に正しいのか後々不安になってきて『2』をつけたが、ステップ3を実行してみて、その情報は間違っていなかったことが分かったので、評価がずれてしまった。最終解決策を行う上で必要な追加情報がうまく設定できなかったので、どこか間違っていると思い『1』をつけた。しかしそれは解決の目標の設定の仕方に問題があったことで、最終解決策は妥当なものだということが分かった」(学生H：自己評価辛め、平均的)のように、改良版トリプルジャンプのステップ1・2の時点では自信がなかったが、後のロールプレイであるステップ3をやりながらその自己評価が修正されていく過程を記述した学生もいた。

以上のような比較から、このような振り返りが、ルーブリックの記述語の意味や課題で求められているものに目を配るきっかけになっていることや、さらに次のステップとして、教員の評価と照らし合わせて、より具体的に自分の成果物をリフレクションし、次の学習に生きるような示唆を得ることを促していると考えられる。

ただし、すべての学生の記述がそのような変化をみせているわけではなく、後期の振り返りにおいても、感覚的に自己評価した旨の記述がみられた。逆に、前期の段

階ですでに、ズレをもとに自身の成果物を再度客観的に分析している学生の記述もみられた。また、学生のサンプルサイズは小さく、記述量も多いものとは言えないため、確定的な結論は保留する必要がある。

しかし、テキストマイニングによって示されたように、全体的な傾向として前期から後期にかけて記述内容の特徴が変化したことから一定の示唆が得られるだろう。まず、前期はルーブリックの解釈に関する言及が目立っていたが、これはSadler (2010) がいう応答ジャンル、基準、質といった評価という技術に関わる概念の理解に関わるものであるといえる。そしてそのような理解を学生が得たあとの後期の評価では、ただルーブリックの記述語の解釈に言及するだけでなく、ズレに対してより分析的な考察ができている記述が見られるようになっている。このような傾向は、振り返りの活動を通して自己評価能力が向上したと解釈する根拠となりえるのではないだろうか。

また、先述の量的分析において、このような振り返りの活動によって、教員による評価を妥当なものとした場合、学生の自己評価がより辛めになるということを明らかにした。その原因として、前期の経験を通して、教員の判断に合わせるために水準を高めに設定し、辛めに自己評価するということが、平均的～平均よりも低い能力の学生にも見られるようになっていることに起因していると解釈できる。ただし、そのような学生の中に、振り返りの中で教員による評価とのズレを認識し、さらに自己評価の修正を行い、自身の良かったところを認識するといった記述をしているものもみられた。

つまり、このような振り返りによって、後期がより辛めの自己評価になった背景には、前期に甘めにつけていたことを認識し、振り返り子のように後期は辛めにつけるようになったということが想定されるが、それをさらに振り返ることで、また自己評価の修正や調整を行っている様子が見て取れた。前期と後期の振り返りによって、繰り返し教員による評価とのズレを認識した彼らが、評価主体として次にどのような自己評価や振り返りをするようになるのか。今後も継続的な調査が必要になるだろう。

#### 4. まとめと今後の課題

本研究は、ルーブリックを学びに活用し、評価主体として育成することを目的とした実践を事例として取り上げた。具体的には、問題解決能力を評価するためのパフォーマンス評価において、ルーブリックを活用し、学生に自己評価させるだけではなく、その評価と教員によ

る評価とのズレを学生に提示し、「なぜズレたのか」の振り返りを促す実践を行い、複数の比較を通してこのような活動が学生の自己評価にどのような影響を与えるのかを検討した。その結果、ループリックの記述語の意味や課題で求められているものを意識するようになった。

また、自分の成果物に対して教員の評価と照らして次の学習へつながるような、より深い考察をするようになるという可能性を見出した。また、評価が相対的に辛くなる傾向を明らかにしたが、同様の振り返りを再び行うことで、さらなる自己評価の修正が促される可能性を見出した。これは、ループリックを学生の学びのために活用する一つの方向性になり得るだろう。

最後に、本研究の限界と今後の課題を述べたい。先述のようにサンプルサイズが小さく、またあくまで特定の歯学系のPBLの授業におけるパフォーマンス評価における実践を事例としたものであるため、結果の再現性や一般化可能性には限界がある。他の評価や他の学問領域の事例の蓄積も肝要であろう。さらに、本研究は半年間という比較的短い期間において、限定的な手法によって自己評価の変容を捉えようとしたものであることに留意しなければならない。今後、継続的な調査によって、再現性の確認や、このような振り返りが長期的にみて学生の自己評価能力の形成にどのような影響を与えたのか、慎重に検討していく必要があるだろう。

また、本研究は教員による評価の数値のみをフィードバックし振り返りを促したものであるが、数値のみではなく、教員がなぜそのような評価をしたのかといったコメントのフィードバックによって、振り返りがより深まり、自己評価能力の形成に有効であると考えられ、そのような検討にも意義があろう。ただし、教員の評価負担を増やすことになるため、実行可能性との兼ね合いも考慮する必要がある。

なお、本研究では教員による評価をより妥当なものとして扱い、そのズレから学生の自己評価の様相を捉えようとした。しかしそれは、教員による評価を絶対的に正確な評価とみなしていると意味するわけではないし、またそのようにみなすことはすでに述べたように危険であることに留意されたい。本研究の実践ではそれを学生にも意識してもらうために、振り返りのプリントでは、教員による評価に納得いかなかった場合に、自己評価が妥当であることを説明するように求めている。何かしらの構成概念を変数化する際、どのような評価手法や測定ツールを用いたとしても、必ず誤差がつきまとう。特にズレや変容などを検討する際には、誤差の影響は顕著になる。当然、そのような誤差が本研究の量的分析に影響

を与えている可能性も考えられる。

また、自己評価能力に関わるメタ認知は、岡本(2012)も述べているように、多面的で曖昧な概念であり、評価や測定が困難な心的過程である。そのため、学生の自己評価能力を捉えることも困難だと考えられる。しかし昨今の教育改革の動向を考慮すると、日々の教育実践のなかで学生の自己評価能力をいかに捉え、いかに育成するののかという議論はますます活発になっていくと予想される。本研究がそのような議論の一助となれば幸いである。

## 謝辞

本研究において、調査に協力してくださった新潟大学歯学部の教員の皆様、学生の皆様に感謝を申し上げます。

また本研究は、JSPS科研費「能力形成を促すパフォーマンス評価の開発と拡張」(平成27～29年度、研究代表者：松下佳代、課題番号15H03473)の助成を受けました。

## 文献

- Fadel, C., Bialik, M., & Trilling, B. (2015). *Four-dimensional education: The competencies learners need to succeed*. Boston, MA: The Center for Curriculum Redesign.
- 樋口耕一 (2014)『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して—』ナカニシヤ出版。
- 伊藤崇達 (2012)「自己調整学習とメタ認知」自己調整学習研究会(編)『自己調整学習 理論と実践の新たな展開へ』, 北大路書房, pp.31-53.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121-1134.
- 松下佳代 (2012)「パフォーマンス評価による学習の質の評価：学習評価の構図の分析にもとづいて」,『京都大学高等教育研究』18, pp.75-114.
- 松下佳代 (2014)「学習成果としての能力とその評価—ループリックを用いた評価の可能性と課題—」,『名古屋高等教育研究』14, pp.235-255.
- 松下佳代 (2016)「資質・能力の新たな枠組み—「3・3・1 モデル」の提案—」,『京都大学高等教育研究』22, pp.139-149.

- 松下佳代・石井英真編 (2016)『アクティブラーニングの評価』東信堂。
- 松下佳代・小野和宏・高橋雄介 (2013)「レポート評価におけるルーブリックの開発とその信頼性の検討」,『大学教育学会誌』35-1, pp.107-115.
- 中山留美子 (2013)「アクティブ・ラーナーを育てる能動的学修の推進におけるPBL教育の意義と導入の工夫」,『21世紀教育フォーラム』8, pp.13-21.
- 岡本真彦 (2012)「教科学習におけるメタ認知」,『教育心理学年報』51, pp.131-142.
- 沖裕貴 (2014)「大学におけるルーブリック評価導入の実際：公平で客観的かつ厳格な成績評価を目指して」,『立命館高等教育研究』14, pp.71-90.
- 小野和宏・松下佳代・斎藤有吾 (2014)「PBLにおける問題解決能力の直接評価—改良版トリプルジャンプの試み—」,『大学教育学会誌』36-1, pp.123-132.
- Rychen, D. S., & Salganik, L. H. (2003). *Key competencies: For a successful life and a well-functioning society*. Boston, MA: Hogrefe & Huber. ライチェン, D. S.・サルガニク, L. H. (2006).『キー・コンピテンシー—国際標準の学力をめざして—』(立田慶裕監訳) 明石書店。
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- 須長一幸 (2010)「アクティブ・ラーニングの諸理解と授業実践への課題—activeness概念を中心に—」,『関西大学高等教育研究』1, pp.1-11.
- 斎藤有吾・小野和宏・松下佳代 (2017)「パフォーマンス評価における教員の評価と学生の自己評価・学生調査との関連」,『日本教育工学会論文誌』40 (Suppl.), pp.157-160.
- 田中耕治 (2008)『教育評価』岩波書店。
- 寺嶋浩介・林朋美 (2006)「ルーブリックの構築により自己評価を促す問題解決学習の開発」,『京都大学高等教育研究』12, pp.63-71.
- 山田勉・斎藤有吾 (2017)「アクティブラーニングの評価への学生参加のあり方」,第23回大学教育研究フォーラム(参加者企画セッション「アクティブラーニングの評価のフロンティア」)発表資料,2017年3月20日。
- 山田嘉徳・森 朋子・毛利美穂・岩崎千晶・田中俊也 (2015)「学びに活用するルーブリックの評価に関する方法論の検討」,『関西大学高等教育研究』6, pp.21-30.
- Zimmerman, B. J. (2001). Models of self-regulated learning and academic achievement. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: theory, research, and practice*. New York: Springer. pp. 1-25.『自己調整学習の理論』(塚野州一編訳) 北大路書房, pp.136.